IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12427

Drug Side Effects Prediction by Using ML And NLP

M.Asha[1], D. Susmitha[2], K. Danamma[3], B. Prema Bhanu[4], B. Venkata Sai Bhavani[5]

M.Tech., Asst.Professor, Computer Science & Engineering,

Bapatla Women's Engineering College, Bapatla, India[1]

B. Tech, Computer Science & Engineering, Bapatla Women's Engineering College, Bapatla, India[2-5]

Abstract: Adverse drug side effects pose significant challenges in pharmaceutical research, ranking as a leading cause of treatment failure and mortality. Traditional laboratory-based evaluations of drug side effects are resource-intensive and time-consuming, necessitating the adoption of machine learning techniques for efficient and accurate predictions. This study explores the use of supervised learning approaches for drug side effect prediction, leveraging biomedical data and computational models. We employ various feature extraction techniques, including Bag of Words (BOW) and Term Frequency-Inverse Document Frequency (TF-IDF), combined with classification models such as Logistic Regression, Random Forest, and Support Vector Machines (SVM). Experimental results demonstrate that the TF-IDF-based models achieve superior performance, with Logistic Regression attaining a test accuracy of 80.88% and SVM achieving 80.90%. These findings highlight the potential of machine learning in predicting drug side effects, optimizing drug safety assessments, and reducing the risks associated with adverse reactions. Our study provides a comprehensive analysis of model effectiveness and discusses key challenges, research gaps, and future directions for improving predictive performance in this critical domain.

Keywords: Machine Learning, Supervised Learning, Feature Extraction, TF-IDF, Bag of Words, Logistic Regression, Random Forest and Support Vector Machines.

I. INTRODUCTION

Drug-related side effects, also known as adverse drug reactions (ADRs), refer to unintended and potentially harmful responses to medications that occur at standard therapeutic doses. These reactions can range from mild discomfort to severe health complications, affecting various organs and tissues. Notably, ADRs are a leading cause of drug development failures and can result in the withdrawal of approved medications from the market. Traditionally, the identification of such side effects has relied on extensive clinical trials, which, while thorough, are both time-consuming and costly. This underscores the necessity for more efficient and cost-effective predictive methods in the pharmaceutical industry.

During the drug development process, various computational techniques have been introduced to predict potential side effects at different stages (refer to reviews by Ho et al. and Boland et al.). The first category of these methods is designed for pre-clinical drug development, where only chemical, biological, and pharmacological data are accessible. These approaches utilize chemical properties, protein targets, and pathway details, often combined with protein networks, to make predictions. However, their accuracy tends to be moderate. Another category of methods focuses on the postmarketing phase, where side effect data collected from clinical trials and real-world usage help predict additional unknown adverse effects. Unlike these methods, our study considers only the side effects observed during clinical trials, making the prediction task more challenging due to limited data and selection bias.

Our research aims to achieve two key objectives: (1) to create realistic scenarios similar to those encountered by safety professionals in clinical drug development and (2) to develop a computational tool that aids in early detection of potential drug side effects during clinical trials. A significant application of our approach lies in different phases of clinical testing, where computational predictions serve as hypothesis generators, guiding risk assessment efforts. We introduce a matrix completion framework named the geometric self-expressive model (GSEM), which employs an optimization function and a multiplicative learning algorithm to find optimal solutions. This model integrates known drug-side effect relationships with structural information derived from chemical, biological, and pharmacological data. Our findings also highlight the difficulty of predicting side effects that emerge after a drug reaches the market based on clinical trial data alone. This is largely due to differences in side effect distribution between clinical trials and postmarketing reports. To address this challenge, we propose a simple data integration technique that significantly enhances GSEM's ability to identify potential postmarketing side effects.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12427

In recent years, the integration of advanced computational algorithms, particularly machine learning techniques, has revolutionized the prediction of drug-related side effects. By analysing extensive datasets that include chemical, biological, and phenotypic information, these models can forecast potential adverse reactions, thereby enhancing patient safety and optimizing drug development processes. Despite these advancements, the drug development pipeline remains fraught with challenges, as approximately 90% of new drugs fail during early human trials, with ADRs contributing to about 35% of these failures. Moreover, the practice of combination therapy, especially among the elderly, increases the risk of drug-drug interactions, accounting for a significant proportion of adverse drug events.

Emphasizing the study and prediction of drug side effects is crucial for improving patient outcomes, refining treatment protocols, and guiding pharmaceutical companies in designing safer medications. By leveraging computational models and machine learning approaches, the industry can better anticipate and mitigate the risks associated with adverse drug reactions, ultimately contributing to the advancement of personalized medicine and overall public health.

II. MATERIALS AND METHODS

Advancements in biomedical technology have accelerated the drug discovery and development process [1]. However, despite these improvements, the process remains lengthy, costly, and high-risk, with nearly 90% of new drugs failing during initial human trials, primarily due to adverse drug reactions, which account for approximately 35% of failures [2]. Adverse drug reactions, commonly known as drug side effects, are officially defined by China's State Food and Drug Administration as "harmful reactions that occur under normal usage and dosage of qualified drugs and have nothing to do with the purpose of medication" [3]. Alarmingly, drug side effects rank as the fourth leading cause of death, following cardiovascular disease, cancer, and infectious diseases [4]. In the mid-20th century, insufficient regulatory oversight led to catastrophic drug-related incidents, resulting in nearly 20,000 deaths and leaving countless others harmed [5]. These adverse effects not only pose significant health risks but also place a financial burden on patients and strain healthcare systems, leading to increased morbidity and mortality rates [6]. Combination therapy, which involves the concurrent use of multiple medications, is often more effective than single-drug treatments and is widely preferred [7]. Many patients, especially the elderly, rely on multiple prescription and over-the-counter medications to manage distinct medical conditions, yet Drug-Drug Interactions (DDIs) contribute to 30% of all adverse drug events and are clinically relevant for up to 80% of elderly cancer patients [8]. Therefore, studying drug side effects is of paramount importance, as it enhances patient safety by identifying and monitoring potential adverse events, thereby reducing treatment risks [9]. Additionally, such research improves treatment protocols, enables healthcare professionals to make informed decisions, enhances patient management, and promotes patient autonomy and awareness regarding their treatments [10]. Furthermore, pharmaceutical companies benefit from these studies as they aid in optimizing drug design and development, ultimately improving the safety and efficacy of medications [11].

In recent years, global efforts have intensified to ensure the safety of widely used pharmaceuticals [12]. Drug safety is a critical consideration throughout the entire lifecycle of drug research and development, with significant implications for public health and overall well-being [13]. During the initial drug design phase, researchers carefully select biomolecules or cellular structures as drug targets to regulate specific physiological processes, thereby reducing the risk of unintended side effects [14]. By ensuring that drugs act selectively on their intended targets without disrupting normal physiological functions, researchers can enhance treatment precision and minimize adverse reactions. Medicinal chemists further refine drug molecules by optimizing their structures to improve selectivity and binding affinity, which lowers the likelihood of off-target effects and associated side effects [15]. As the drug development process advances, extensive studies on drug metabolism and toxicity are conducted to analyze metabolic pathways and identify potential risks [16]. These studies play a crucial role in predicting and mitigating adverse reactions, though they require complex analyses of molecular structures and Structure-Activity Relationships (SARs), as well as costly and time-intensive pharmacological risk assessments [17]. Clinical trials are then conducted to assess drug efficacy and safety, allowing researchers to adjust dosages, modify treatment durations, or determine whether to proceed with further development based on observed side effects [18]. However, even with rigorous pre-approval testing, identifying all possible adverse drug reactions remains challenging due to genetic diversity, metabolic variability, and the limitations of clinical trial sample sizes [19]. As a result, post-marketing surveillance has become an essential component of drug safety assessment [20]. This phase relies on public databases where healthcare professionals report observed side effects, but since reporting is voluntary, some adverse reactions may go undocumented, limiting the accuracy of safety evaluations. Furthermore, drug side effects arise through highly intricate mechanisms. Some side effects occur when drugs unintentionally interact with proteins structurally similar to their intended targets, triggering unintended biological responses. Additionally, metabolic processes can produce drug metabolites with varying toxicity levels, and individual sensitivity to these metabolites can lead to diverse adverse reactions.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

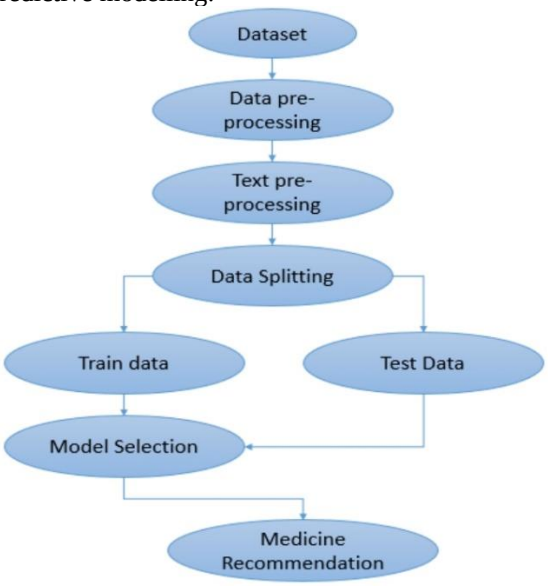
DOI: 10.17148/IARJSET.2025.12427

Several factors, including immune system responses, drug interactions, genetic predispositions, drug concentration levels, and individual physiological differences, further contribute to the complexity of drug side effect formation. Given these challenges, the development of reliable computational models for predicting drug side effects has become a critical area of research, aiming to enhance drug safety and reduce risks associated with adverse reactions.

The integration of machine learning techniques into scientific research has gained significant momentum in recent years. These methods not only facilitate the efficient processing of large datasets but also help uncover patterns, trends, and even predict unresolved challenges. Their versatility has led to applications across various fields, including medicine, biology, engineering, and finance [21]. Utilizing machine learning for analyzing drug side effects offers several advantages. Firstly, it enables the consolidation of vast amounts of medical data from multiple sources, providing a comprehensive understanding of adverse effects, including rare and newly emerging ones. Secondly, machine learning algorithms automatically identify patterns and correlations within the data, assisting in the early detection of adverse effects that might otherwise remain unnoticed. Additionally, these methods can predict potential side effects and identify patients at higher risk, allowing for proactive interventions to enhance patient safety. The growing body of literature on drug side effects, along with the expansion of online databases containing extensive information on drugs, adverse reactions, and related biological factors, provides researchers with abundant opportunities to refine machine learning models for side effect prediction. While previous reviews on drug side effect prediction have primarily focused on methods for identifying drug-side effect associations, this study presents a more comprehensive perspective by categorizing machine learning-based techniques for predicting side effects caused by both individual drugs and Drug-Drug Interactions (DDIs). Furthermore, it highlights approaches for estimating the frequency and severity of side effects. The methodologies are systematically classified into four distinct categories based on specific prediction tasks. The structure of this paper is as follows: Section 2 explores various datasets and web resources relevant to drug side effect prediction, while Sections 3 and 4 examine different machine learning techniques designed for this purpose. Section 5 discusses current challenges and future research directions in drug side effect prediction, and Section 6 concludes the study by outlining potential advancements and promising trends in this field.

III. METHODOLOGY

The proposed architecture for drug side effect prediction follows a structured pipeline comprising multiple stages: data collection, preprocessing, feature extraction, model training, and evaluation. Initially, data is sourced from publicly available biomedical datasets on platforms such as Kaggle, containing a mix of categorical, numerical, and textual attributes related to drug properties and adverse reactions. The preprocessing phase begins with handling missing values using statistical imputation—mean and median for numerical attributes and mode for categorical ones. Textual data is preprocessed through tokenization, stopword removal, and stemming or lemmatization to enhance text quality. To prepare data for machine learning models, categorical features are encoded using Label Encoding (LE), while numerical attributes are standardized using Standard Scaling (SS). Additionally, textual features are converted into numerical representations using Bag of Words (BOW) and Term Frequency-Inverse Document Frequency (TF-IDF), ensuring meaningful pattern extraction for predictive modelling.



© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12427

Once the data is preprocessed and transformed, multiple supervised learning models are trained to classify potential drug side effects. The selected models include Logistic Regression, Random Forest, and Support Vector Machines (SVM), each trained separately using both BOW and TF-IDF representations to compare their effectiveness. The models undergo rigorous evaluation using standard performance metrics such as precision, recall, and F1-score to determine their ability to predict adverse drug reactions accurately. By leveraging machine learning techniques, this approach aims to enhance drug safety assessments by identifying potential risks early in the drug development pipeline. Future extensions of this work may incorporate deep learning architectures or additional biomedical knowledge sources to improve predictive accuracy and expand the scope of side effect analysis.

Dataset: - The dataset used for this study is sourced from the **KUC Hackathon - Winter 2018** on Kaggle, which contains a diverse set of features representing drug characteristics and potential adverse reactions. This dataset includes a mix of **categorical, numerical, and textual attributes**, making it suitable for applying machine learning techniques for predictive analysis. The categorical features typically represent drug classifications, side effect categories, and other qualitative properties, while numerical features provide dosage information and other quantitative measures. Additionally, the dataset contains textual data that describes drug properties and effects, requiring preprocessing techniques such as tokenization and feature extraction. Since real-world medical data often contains inconsistencies, handling missing values, standardizing numerical attributes, encoding categorical features, and transforming text into numerical representations are crucial preprocessing steps. This dataset serves as a valuable resource for training predictive models, allowing for the identification of patterns in drug-related side effects and supporting the development of safer medications.

Data preprocessing:

Data Preprocessing is a vital step to ensure the dataset is clean and structured for effective machine learning modeling. The process begins with **removing duplicate entries**, as repeated records can introduce bias and redundancy, negatively affecting model performance. Next, we handle **missing values**, which are common in real-world datasets. For **numerical features**, missing values are imputed using the **mean** if the data follows a normal distribution or the **median** for skewed distributions. This prevents data loss while maintaining statistical integrity. For **categorical features**, missing values are replaced with the **most frequent category**, ensuring that important patterns in the data remain intact. If an entire row contains only missing values, it is removed from the dataset to avoid unnecessary noise.

Following this, **Exploratory Data Analysis (EDA)** is conducted to detect **outliers**, which can significantly impact model predictions. Outliers in numerical features are identified using **box plots, z-score analysis, and the interquartile range (IQR) method**. Depending on their severity, they are either removed or adjusted through capping techniques. For categorical features, we analyze frequency distributions to identify rare or inconsistently represented categories, which may need grouping or elimination. Additionally, **scaling techniques** like **StandardScaler (SS)** are applied to numerical features to ensure uniformity, while **Label Encoding (LE)** is used to convert categorical variables into numerical form for seamless model processing. These steps collectively ensure that the dataset is refined, structured, and ready for feature extraction and model development.

Text Preprocessing:

Text preprocessing is a crucial step in natural language processing (NLP) to convert raw text into a structured format that machine learning models can effectively analyze. This process involves multiple steps to clean and standardize text data, ensuring that important linguistic features are retained while irrelevant noise is removed.

The first step is **removing punctuation marks**, such as commas, periods, question marks, and special characters. Punctuation does not contribute meaningful information to the predictive model and can introduce unnecessary complexity. By eliminating these symbols, we ensure that only relevant words are processed.

Next, we remove **stop words**, which are common words like "the," "is," "and," and "in" that do not add significant meaning to text analysis. Stop words are filtered out using predefined lists from NLP libraries such as **NLTK** or **spaCy** to reduce the dimensionality of the dataset and improve model efficiency.

Once stop words are removed, we convert all text to **lowercase** to maintain uniformity. Text data often contains words in mixed cases, such as "Drug" and "drug," which are treated as separate entities by machine learning models. Converting all words to lowercase ensures consistency and prevents duplication of features.

The next step is **stemming**, which reduces words to their root form. For example, words like "running," "runs," and "ran" are converted to "run." This is done using algorithms like **Porter Stemmer** or **Snowball Stemmer**, which trim suffixes to maintain a consistent word base. While stemming is effective, it may sometimes produce words that are not linguistically meaningful.

An alternative to stemming is **lemmatization**, which reduces words to their base form while preserving their meaning. Unlike stemming, lemmatization considers the word's context and grammar. For instance, "better" is reduced to "good," and "running" is converted to "run." This is achieved using tools like **WordNetLemmatizer** from the **NLTK** library.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12427

After preprocessing the text, we apply **feature extraction techniques** to convert the cleaned text into a numerical format suitable for machine learning models. The two most common methods are **Bag of Words (BOW)** and **Term Frequency-Inverse Document Frequency (TF-IDF)**.

- **Bag of Words (BOW)**: This technique represents text data as a matrix of word occurrences, where each word in the corpus is assigned a unique index, and the text is converted into a vector containing the frequency of each word. Although simple and effective, BOW does not capture the meaning or relationships between words.
- **Term Frequency-Inverse Document Frequency (TF-IDF)**: Unlike BOW, TF-IDF assigns weights to words based on their importance in a document relative to the entire dataset. It calculates **term frequency (TF)**—the number of times a word appears in a document—and **inverse document frequency (IDF)**—which reduces the weight of commonly occurring words across multiple documents. This technique helps prioritize meaningful words while minimizing the influence of frequently used but less informative terms.

By applying these text preprocessing steps, we ensure that the text data is clean, structured, and ready for machine learning models to extract meaningful insights effectively.

Data splitting: - In machine learning, splitting the dataset into different subsets is crucial for training, validating, and testing the model. For this project, we divide the dataset into three parts: **60% for training, 20% for validation, and 20% for testing**.

- **Training Set (60%)**: The majority of the data is allocated for training the model. This subset helps the model learn patterns and relationships between features and target labels. The model is optimized by adjusting its parameters based on this data.
- **Validation Set (20%)**: After training, the model's performance is evaluated using the validation set. This helps in fine-tuning hyperparameters and preventing overfitting. Since this data is unseen during training, it ensures the model generalizes well.
- **Test Set (20%)**: Once the model is fine-tuned, the final evaluation is done on the test set. This independent dataset provides an unbiased assessment of how the model will perform on real-world, unseen data.

Model Building:

Logistic Regression :- Logistic Regression is a supervised learning algorithm used for classification tasks. It works by applying the sigmoid function to map input features to a probability score between 0 and 1. Based on a defined threshold (typically 0.5), the model classifies data points into different categories. It uses a weighted sum of input features and applies the logistic (sigmoid) function to determine the likelihood of belonging to a particular class. The algorithm optimizes weights using techniques like Gradient Descent to minimize errors and improve predictions. Logistic Regression is efficient, interpretable, and performs well on linearly separable data.

Random Forest :- Random Forest is an ensemble learning algorithm that builds multiple decision trees and combines their outputs to improve classification accuracy. It operates by randomly selecting subsets of data and features to create diverse decision trees, reducing overfitting. Each tree independently makes a prediction, and the final classification is determined by majority voting. This approach enhances model stability, handles non-linearity, and works well with both numerical and categorical data. Additionally, Random Forest provides feature importance rankings, making it useful for understanding which variables influence predictions the most.

Linear Support Vector Machine (SVM) is a supervised learning algorithm used for classification tasks by finding the optimal hyperplane that best separates data points into different categories. It works by maximizing the margin between classes, ensuring better generalization to unseen data. Linear SVM is particularly effective for high-dimensional datasets and is computationally efficient. It performs well when the data is linearly separable, making it a reliable choice for text classification and other machine learning applications.

RESULTS AND DISCUSSION

The performance of various machine learning models for drug side effect prediction was evaluated using different feature extraction techniques, including Bag of Words (BOW) and Term Frequency-Inverse Document Frequency (TF-IDF). The results, presented in Table 1, indicate that models trained using TF-IDF feature representation demonstrated slightly better test accuracy compared to those using BOW. Among all models, TF-IDF combined with Support Vector Machine (SVM) achieved the highest test accuracy of 80.90%, followed closely by Logistic Regression with TF-IDF at 80.88%. While Random Forest exhibited the highest training accuracy across both feature representations, its test performance did not surpass that of TF-IDF-based models. This suggests that Random Forest might be overfitting the training data, leading to a slight reduction in its generalization capability.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12427

Table 1: Model Performance on Drug Side Effect Prediction

Featurization	Model	Train Accuracy	Test Accuracy
BOW	Logistic Regression	0.9427	0.7973
BOW	Random Forest	0.9801	0.7983
BOW	Linear SVM	0.8748	0.8015
TF-IDF	Random Forest	0.9983	0.7983
TF-IDF	Logistic Regression	0.9838	0.8088
TF-IDF	Linear SVM	0.9154	0.809

To further analyze the model performance, we visualized the accuracy distribution using a count plot (Figure 1), which illustrates the training and test accuracy of all models. The results reveal that models utilizing TF-IDF consistently performed better in terms of test accuracy compared to those using BOW. Additionally, Logistic Regression and SVM models demonstrated better generalization, while Random Forest exhibited a notable disparity between training and test accuracy, indicating possible overfitting. These findings emphasize the importance of feature extraction techniques in influencing model performance, where TF-IDF proved to be more effective in capturing relevant text-based features compared to BOW.

Despite achieving promising results, certain challenges remain, including the slight overfitting observed in Random Forest and the need for further optimization of hyperparameters to enhance model performance. Future work could focus on exploring advanced deep learning-based approaches, such as transformer models, to further improve prediction accuracy. Additionally, incorporating domain-specific embeddings could refine the feature representation, leading to more accurate and robust predictions. The findings from this study highlight the potential of machine learning techniques in drug side effect prediction, offering valuable insights for enhancing drug safety assessments.

IV. CONCLUSION

This study demonstrates the effectiveness of machine learning techniques in predicting adverse drug side effects using textual and categorical data. By employing feature extraction methods such as Bag of Words (BOW) and Term Frequency-Inverse Document Frequency (TF-IDF), we evaluated multiple classification models, including Logistic Regression, Random Forest, and Support Vector Machines (SVM).

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12427

The experimental results indicate that TF-IDF-based models outperform BOW-based models, with TF-IDF SVM achieving the highest test accuracy of **80.90%**, followed closely by TF-IDF Logistic Regression at **80.88%**. The Random Forest model also showed competitive performance across both feature extraction techniques. These findings highlight the potential of leveraging machine learning methodologies for improving drug safety assessments and early detection of adverse reactions.

While this study provides valuable insights into drug side effect prediction, several areas require further exploration. First, integrating external biomedical databases, such as FAERS (FDA Adverse Event Reporting System) and DrugBank, could enhance model accuracy by providing richer contextual information. Second, deep learning models, such as transformers and recurrent neural networks (RNNs), could be investigated to capture more complex relationships in drug-related text data. Additionally, expanding the dataset to include multi-modal data, such as molecular structures and patient demographics, could improve prediction robustness. Future research should also focus on explainability techniques to enhance model interpretability, allowing healthcare professionals to trust and utilize AI-driven drug safety assessments more effectively. By advancing these areas, we can further improve the reliability and applicability of machine learning models in pharmacovigilance and personalized medicine.

REFERENCES

[1]. Berdigaliyev N, Aljofan M. An overview of drug discovery and development. Future Medicinal Chemistry, 2020, 12(10): 939−94

[2]. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharmaceutica Sinica B, 2022, 12(7): 3049−3062

[3]. Aronson J K. Adverse drug reactions: history, terminology, classification, causality, frequency, preventability. In: Talbot J, Aronson J K, eds. Stephens' Detection and Evaluation of Adverse Drug Reactions: Principles and Practice. 6th ed. Chichester: John Wiley & Sons, Ltd, 2011, 1−119

[4]. Church D L. Major factors affecting the emergence and re-emergence of infectious diseases. Clinics in Laboratory Medicine, 2004, 24(3): 559−586

[5]. Zullo A, Large M, Amoros E, Martin J L. Estimated number of seriously injured road users admitted to hospital in France between 2010 and 2017, based on medico-administrative data. BMC Public Health, 2021, 21: 649

[6]. De Kinderen R J A, Evers S M A A, Rinkens R, Postulart D, Vader C I, Majoie M H J M, Aldenkamp A P. Side-effects of antiepileptic drugs: the economic burden. Seizure, 2014, 23(3): 184−190

[7]. Mokhtari R B, Homayouni T S, Baluch N, Morgatskaya E, Kumar S, Das B, Yeger H. Combination therapy in combating cancer. Oncotarget, 2017, 8(23): 38022−38043

[8]. Rao A, Cohen H J. Symptom management in the elderly cancer patient: fatigue, pain, and depression. JNCI Monographs, 2004, 2004(32): 150−157

[9]. Forster A J, Worthington J R, Hawken S, Bourke M, Rubens F, Shojania K, Van Walraven C. Using prospective clinical surveillance to identify adverse events in hospital. BMJ Quality & Safety, 2011, 20(9): 756−763

[10]. Papaioannou D, Cooper C, Mooney C, Glover R, Coates E. Adverse event recording failed to reflect potential harms: a review of trial protocols of behavioral, lifestyle and psychological therapy interventions. Journal of Clinical Epidemiology, 2021, 136: 64−76 Bender R, Beckmann L, Lange S. Biometrical issues in the analysis of adverse events within the benefit assessment of drugs. Pharmaceutical Statistics, 2016, 15(4): 292−296

[11]. Rácz A, Bajusz D, Miranda-Quintana R A, Héberger K. Machine learning models for classification tasks related to drug safety. Molecular Diversity, 2021, 25(3): 1409−1424

[12]. Call K T, Riedel A A, Hein K, McLoyd V, Petersen A, Kipke M. Adolescent health and well-being in the twenty-first century: a global perspective. Journal of Research on Adolescence, 2002, 12(1): 69−98

[13]. Eisenhauer E A, O'Dwyer P J, Christian M, Humphrey J S. Phase I clinical trial design in cancer drug development. Journal of Clinical Oncology, 2000, 18(3): 684−684

[14]. Fliri A F, Loging W T, Thadeio P F, Volkmann R A. Analysis of drug induced effect patterns to link structure and side effects of medicines. Nature Chemical Biology, 2005, 1(7): 389−397

BIOGRAPHY

Mrs.M. Asha, M. Tech, Asst. Professor,
Dept of Computer Science & Engineering,
BWEC, Andhra Pradesh, India.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 4, April 2025

DOI: 10.17148/IARJSET.2025.12427

D.Susmitha [B.Tech], Student,
Dept of computer science & Engineering,
BWEC, Andhra Pradesh, India

K. Danamma [B.Tech], Student,
Dept of computer science & Engineering,
BWEC, Andhra Pradesh, India

B. Prema Bhanu[B.Tech], Student,
Dept of computer science & Engineering,
BWEC, Andhra Pradesh, India

B. Venkata Sai Bhavani [B.Tech], Student,
Dept of computer science & Engineering,
BWEC, Andhra Pradesh, India

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License